

AD-A013 085

AN EXPLORATORY STUDY OF THE APPLICATION OF GENERALIZED
INVERSE TO ILS ESTIMATION OF OVERIDENTIFIED EQUATIONS
IN LINEAR MODELS

J. Daniel Khazzoom

Stanford University

Prepared for:

Office of Naval Research

15 April 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 219	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Exploratory Study of the Application of Generalized Inverse to ILS Estimation of Overidentified Equations in Linear Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report
7. AUTHOR(s) J. Daniel Khazzoom		6. PERFORMING ORG. REPORT NUMBER
8. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, Calif. 94305		9. CONTRACT OR GRANT NUMBER(s) N00014-67-A-0112-0085
10. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		11. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-267)
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE April 15, 1975
		12. NUMBER OF PAGES X 18
		13. SECURITY CLASS. (of this report) Unclassified
		13a. DECLASSIFICATION/DOWNGRADING SCHEDULE
14. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release: Distribution Unlimited.		
15. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
16. SUPPLEMENTARY NOTES		
17. KEY WORDS (Continue on reverse side if necessary and identify by block number) Econometric model; overidentification; Moore-Penrose generalized inverse; Indirect Least Squares; Two Stage Least Squares; Instrumental Variables; Limited Information Maximum Likelihood; Full Information Maximum Likelihood		
18. ABSTRACT (Continue on reverse side if necessary and identify by block number) In this paper, we propose a procedure based on the use of the Moore-Penrose inverse of matrices for deriving unique Indirect Least Squares (ILS) esti- mates of the structural parameters in the overidentified case. The procedure makes use of all reduced form estimates in deriving the unique structural estimates. The estimator is shown to be consistent. We derive the relation- ship between this Two Stage Least Squares (2SLS) estimator and Instrumental- Variables (I.V.) estimators. We also derive the asymptotic distribution of the proposed estimator. The results of sampling experiments are summarized.		

DD FORM 1473
1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-01-11 5011

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AN EXPLORATORY STUDY OF THE APPLICATION OF GENERALIZED INVERSE
TO ILS ESTIMATION OF OVERIDENTIFIED EQUATIONS IN LINEAR MODELS

by

J. Daniel Khazzoom

Technical Report No. 219
April 15, 1975

Prepared under Contract N00014-67-A-0112-0085
(NR-042-267)
Office of Naval Research
Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ia

AN EXPLORATORY STUDY OF THE APPLICATION OF GENERALIZED INVERSE
TO ILS ESTIMATION OF OVERIDENTIFIED EQUATIONS IN LINEAR MODELS

by

J. Daniel Khazzoom

Abstract

In this paper, we propose a procedure based on the use of the Moore-Penrose inverse of matrices for deriving unique Indirect Least Squares (ILS) estimates of the structural parameters in the overidentified case. The procedure makes use of all reduced form estimates in deriving the unique structural estimates. The estimator is shown to be consistent. We derive the relationship between this Two-Stage Least Squares (2SLS) estimator and Instrumental-Variables (I.V.) estimators. We also derive the asymptotic distribution of the proposed estimator. The results of sampling experiments are summarized.

1. ILS (Indirect Least Squares) Set-up

Let the operation of an economic system be characterized by

$$(1.1) \quad \begin{matrix} Y \\ (Txm) \end{matrix} = \begin{matrix} Y \\ (Txm) \end{matrix} \begin{matrix} B \\ (m \times n) \end{matrix} + \begin{matrix} X \\ (TxG) \end{matrix} \begin{matrix} C \\ (Gxm) \end{matrix} + \begin{matrix} U \\ (Txm) \end{matrix}.$$

Y , X , and U are matrices of endogenous, predetermined and random variables, respectively; B and C are parameter matrices of β 's and γ 's, respectively. (Throughout, we follow essentially the notations synthesized by Dhrymes [3, pp. 172-200, 279-365] and Rao and Mitra [8, pp. 12-17].)

The dependent variable y_{ti} is explained by $m_1 < m$ current endogenous variables and $G_1 \leq G$ predetermined variables. We make the usual assumptions on the random matrix U . The reduced form of (1.1) is defined as

$$(1.2) \quad Y = XC(I-B)^{-1} + U(I-B)^{-1} = X\Pi + V.$$

By appropriately partitioning $\hat{\Pi}$ ($\hat{\cdot}$ indicates estimates), postmultiplying $\hat{\Pi}$ by the $m \times 1$ column $[1, -\hat{\beta}_{.1}, 0]'$ and rearranging terms, we have the usual recursive system for the estimated parameters of the first equation in (1.2):

$$(1.3) \quad \begin{bmatrix} \hat{\Pi}_{G_1 m_1} & I_{G_1} \\ \hat{\Pi}_{G^* m_1} & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_{.1} \\ \hat{\gamma}_{.1} \end{bmatrix} = \begin{bmatrix} \hat{\Pi}_{G_1 1} \\ \hat{\Pi}_{G^* 1} \end{bmatrix}.$$

It is well known that (1.3) has a unique solution in the just-identified case (that is, when $\hat{\Pi}_{G^* m_1}$ is non-singular). In the overidentified case ($G^* > m_1$, and $\hat{\Pi}_{G^* m_1}$ has full column rank) there is more than one way (although a finite number of ways) for consistently estimating the structural parameters.

Because of the difficulty in choosing among these alternative ways, ILS

has fallen into practical disuse. Other limited information estimators which were developed in the meantime compromise in various ways between the estimates in the overidentified case. The procedure I propose in this paper is a compromise in the same nature as the 2SLS. We use the Moore-Penrose (MP) generalized inverse to derive unique ILS estimates of the structural parameters in the overidentified case. For a discussion of the MP inverse, see [8, pp. 50-55]. Briefly, if D is an $m \times n$ matrix, its MP inverse, denoted by D^+ , is an $n \times m$ matrix which satisfies the following four conditions: i) $DD^+D = D$; ii) $D^+DD^+ = D^+$; iii) $(DD^+)' = DD^+$; iv) $(D^+D)' = D^+D$ where $'$ denotes conjugate transpose and where the inner product is defined with respect to the identity matrix. D^+ is unique and has the same rank as D . A matrix that satisfies (i) only is called a g -inverse and usually denoted by D^- .¹

2. ILS Estimates Using Moore-Penrose Inverse

Denote equation (1.3) as

$$(2.1) \quad \hat{D}\hat{\delta}_{.1} = \hat{\pi}_{.1}.$$

Since D has full column rank, $\hat{D}^+ = (\hat{D}'\hat{D})^{-1}\hat{D}'$. If we use the MP inverse to solve (2.1), we get

$$(2.2) \quad \hat{\delta}_{.1} = (\hat{D}'\hat{D})^{-1}\hat{D}'\hat{\pi}_{.1}.$$

The vector $\hat{\delta}_{.1}$ is unique and has the property that it is minimum (Euclidean) norm least squares solution of (2.1). Using the fact that $\hat{D}^+ = (\hat{D}'\hat{D})^{-1}\hat{D}'$, it follows that

$$\hat{D}^+ = \begin{bmatrix} 0 & (\hat{\pi}_{G^*m_1})^+ \\ I & -\hat{\pi}_{G_1m_1}(\hat{\pi}_{G^*m_1})^+ \end{bmatrix},$$

which shows that $\hat{\delta}_{.1}$ in (2.2) coincides with the recursive solution of (1.3), when the MP inverse is used to solve for $\hat{\beta}_{.1}$ and then $\hat{\gamma}_{.1}$.

It can be shown that $\hat{\delta}_{.1}$ is a consistent estimator of $\delta_{.1}$. To see this, note that the elements of \hat{D} are consistent estimates of the corresponding elements in the true D . Since $D'D$ is non-singular, it follows that

$$(2.3) \quad \text{plim } \hat{\delta}_{.1} = (D'D)^{-1} D' \pi_{.1} = D^+ \pi_{.1} = \delta_{.1}.$$

The last equality follows from the fact that in the population the system $D\delta_{.1} = \pi_{.1}$ is known to be a consistent set of equations.

3. Relation of ILS to 2SLS and I.V. Estimator; Asymptotic Distribution

Writing in full the first equation in (1.1), we have

$$(3.1) \quad y_{.1} = Y_1 \beta_{.1} + X_1 \gamma_{.1} + u_{.1} = X\hat{\pi}_{.1} \beta_{.1} + X_1 \gamma_{.1} + u_{.1} + \hat{v}_{.1} \beta_{.1}.$$

$\hat{\pi}_{.1}$, $\hat{v}_{.1}$ consist of the 2nd, ..., (m_1+1) st column of $\hat{\Pi}$, and \hat{v} , respectively. Noting

$$(3.2) \quad (X\hat{\pi}_{.1}, X_1) = X\hat{D},$$

the 2SLS solution of (3.1) is easily seen to be

$$(3.3) \quad \tilde{\delta}_{.1} = \{(X\hat{D})'(X\hat{D})\}^{-1} (X\hat{D})' y_{.1} = \{\hat{D}'(X'X)\hat{D}\}^{-1} \hat{D}'(X'X)\hat{\pi}_{.1},$$

where we made use of the fact that $y_{.1} = X\hat{\pi}_{.1} + \hat{v}_{.1}$ and $X'\hat{v}_{.1} = 0$.

Observe (3.3) is the minimum norm least squares solution of

$$(3.4) \quad X\hat{D}\tilde{\delta}_{.1} = X\hat{\pi}_{.1}.$$

By comparing (2.2) with (3.3) it is evident how the 2SLS and ILS compromise between the various estimates in the overidentified case.

Both procedures solve for the minimum norm least squares estimator. The difference is in the definition of the norm. In the 2SLS case, the (quadratic) norm in (3.3) is defined with respect to the moment matrix of all the predetermined variables. In the ILS case, the norm is defined with respect to the identity matrix. The two estimators coincide when $X'X$ is a scalar matrix, and similarly when the equation is just identified (\hat{D} is then square and non-singular). In light of (2.2) and (3.3), it is straightforward to infer the asymptotic distribution of the proposed ILS estimator from the asymptotic distribution of the 2SLS estimator. For 2SLS, we have (see Dhrymes [3, pp. 190-192])

$$(3.5) \quad \sqrt{T} (\hat{\delta}_{.1} - \delta_{.1}) \sim N(0, \sigma_{11} \text{ plim } \phi_t),$$

where

$$(3.6) \quad \phi_t = \left[\frac{X'Z_1}{T} \right]^+ \left[\frac{X'X}{T} \right]^{-1} \cdot \left[\frac{X'X}{T} \right] \cdot \left[\frac{Z_1'X}{T} \right]^+ \left[\frac{X'X}{T} \right]^{-1},$$

$$(3.7) \quad Z_1 = (Y_1 \ X_1).$$

The subscript $\left[\frac{X'X}{T} \right]^{-1}$ indicates the matrix with respect to which the inner product is defined. For ILS, we have

$$(3.8) \quad \sqrt{T} (\hat{\delta}_{.1} - \delta_{.1}) \sim N(0, \sigma_{11} \text{ plim } \psi_t),$$

$$(3.9) \quad \psi_t = \left[\frac{X'Z_1}{T} \right]^+ \left[\frac{X'X}{T} \right]^{-2} \cdot \left[\frac{X'X}{T} \right] \cdot \left[\frac{Z_1'X}{T} \right]^+ \left[\frac{X'X}{T} \right]^{-2}.$$

In order to arrive at ψ_t we simply changed the norm in (3.6) so that inner product of \hat{D}^+ is now defined with respect to the identity matrix rather than $(X'X)$. Equation (3.9) can also be derived directly.

It is also useful to look at the results derived so far from the point of view of I.V. estimation. If we choose the instrumental variables

$$(3.10) \quad \begin{aligned} P &= X(X'X)^{-1}X'Z_1, \\ Q &= X(X'X)^{-1}, \end{aligned}$$

and rewrite (3.6) and (3.9) as

$$(3.11) \quad \phi_t = \left[\frac{P'Z_1}{T} \right]^{-1} \frac{P'P}{T} \left[\frac{Z_1'P}{T} \right]^{-1},$$

$$(3.12) \quad \psi_t = \left[\frac{Q'Z_1}{T} \right]^+ \frac{Q'Q}{T} \left[\frac{Z_1'Q}{T} \right]^+,$$

we see that the righthand side of (3.11) and (3.12) has the standard form of the covariance matrix whose probability limit appears in the asymptotic distribution of I.V. estimators (except that in (3.12) we have MP inverse instead of the conventional inverse). This is not a surprising result, since it is well known that 2SLS and ILS have an I.V. interpretation, with instruments P and Q , respectively, as in (3.10). Where (3.12) departs from conventional results, however, is in the number of instrumental variables. Q has G columns where $G > m_1 + G_1$, whereas it is standard to require the number of instrumental variables to be the same as the number of explanatory variables in the equation (otherwise the matrix to be inverted will not be square in the first place). Dhrymes [3, p. 365], for example, points out that when $G > m_1 + G_1$, the ILS will fail to yield unique estimators because of what may be interpreted as the attempt to use "too many" instrumental variables in estimating the structural parameters. The results we derived in this section indicate that "too many" instruments is not really a hindrance for deriving unique I.V.

estimates, if we are willing to work with the MP inverse. (Note that if instead of P as defined in (3.10), we choose $P = X(X'X)^{-1}X'$ and substitute in (3.14) below for this alternative choice of P , we would get (3.4), which we already know yields the 2SLS estimator when MP inverse is used to solve it).

In an asymptotic efficiency sense, the 2SLS dominates the class of I.V. estimators whose instruments belong to the subspace spanned by the predetermined variables of the system. For the two covariance matrices (3.11) and (3.12), the relative efficiency (deleting the division of T),

$$(3.13) \quad \psi_t - \phi_t = \hat{D}^+ Q' [I - P(P'P)^{-1}P'] Q \hat{D}^+,$$

is a p.s.d. matrix, since $I - P(P'P)^{-1}P'$ is symmetric idempotent. To summarize: the ILS estimator proposed in this paper, as well as the 2SLS estimator in the overidentified case, achieves a compromise among the various estimates in the overidentified case by finding the (unique) minimum norm least squares solution for $\delta_{.1}$ in

$$(3.14) \quad P'y_{.1} = P'Z_1 \delta_{.1},$$

$$(3.15) \quad Q'y_{.1} = Q'Z_1 \delta_{.1},$$

where P and Q are defined in (3.10). The solution of (3.14) yields the 2SLS estimator and the solution of (3.15) yields the ILS estimator. In both cases, the solutions have the same structure; they differ in the definition of the norm.

4. Design of the Monte Carlo Experiments

The objective of the experiments is (1) to gather evidence on the relative performance of ILS vs. 2SLS and Limited Information Maximum Likelihood (LIML) estimator--the two most commonly used single-equation consistent estimators; (2) to test the hypothesis that the bias for the ILS estimator does not depend on (a) the size of the covariance matrix $\Sigma = (\sigma_{ij})$ of the vector $(u_{t1}, u_{t2}, \dots, u_{tm})'$, (b) the sparseness of Σ , and (c) the sample size; (3) to test the hypotheses that the relative performance of ILS does not depend on the factors listed in (a) to (c). For the purpose of this paper, I chose a structure from one of the experiments reported by Cragg [2]. (Initially, I estimated several runs using the structure estimated by Wagner [11], but because of the special nature of the structure used by Wagner--damped difference equations dominated by a trend factor--I did not think the results would be of general interest.) The rationale for using this particular structure was to permit a comparison of ILS with 2SLS and LIML in a model for which the last two estimators are known to have performed very well. The structure is the following:

$$\begin{aligned} y_{1t} &= .89y_{2t} + .16y_{3t} + 44.00x_{1t} + .74x_{2t} + .13x_{3t} + u_{1t} \\ y_{2t} &= .74y_{1t} + 62.00x_{1t} + .96x_{3t} + .70x_{5t} + .06x_{7t} + u_{2t} \\ y_{3t} &= .29y_{2t} + 40.00x_{1t} + .11x_{4t} + .53x_{5t} + .56x_{6t} + u_{3t} \end{aligned}$$

where x_{1t} is a vector of 1's. In conjunction with Σ_1 (see below), this is structure 8 reported in Cragg [2, p. 92]. In all experiments, we estimated the parameters of the first equation only, using ILS, 2SLS and LIML. An experiment consisted of generating 100 samples of Size $T = 60$,

40, and 20 observations. The predetermined variables are truly exogenous and, except for the vector of constants x_1 , are uniformly and independently distributed random numbers with values in $[-100,100]$. The values of the exogenous variables were fixed for repeated samples of the same size. The sample correlation matrices for the exogenous variables used in the experiments are the following:

$$\begin{array}{ccc}
 T=60 & T=40 & T=20 \\
 \begin{array}{c} x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \\
 \begin{bmatrix} .23 \\ .10 \quad -.05 \\ .02 \quad .29 \quad -.02 \\ .02 \quad .03 \quad .01 \quad .02 \\ .03 \quad -.13 \quad -.10 \quad .05 \quad -.13 \end{bmatrix}
 \end{array} &
 \begin{array}{c} x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \\
 \begin{bmatrix} .16 \\ -.04 \quad -.33 \\ .45 \quad -.08 \quad -.16 \\ .29 \quad .02 \quad .15 \quad .19 \\ .00 \quad -.04 \quad .19 \quad -.07 \quad -.02 \end{bmatrix}
 \end{array} &
 \begin{array}{c} x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \\
 \begin{bmatrix} .01 \\ -.02 \quad -.24 \\ .51 \quad .33 \quad -.19 \\ .09 \quad .07 \quad -.34 \quad .15 \\ -.01 \quad -.19 \quad .18 \quad -.43 \quad -.43 \end{bmatrix}
 \end{array}
 \end{array}$$

The structural disturbances were generated from mutually independent and normally distributed (3-dimensional) vectors with zero mean and the following Σ 's:

$$\begin{array}{cc}
 \Sigma_1 = \begin{bmatrix} 38.60 \\ -5.92 \quad 36.68 \\ -14.80 \quad -2.98 \quad 40.64 \end{bmatrix}; & \Sigma_1^0 = \begin{bmatrix} 38.60 \\ 0 \quad 36.68 \\ 0 \quad 0 \quad 40.64 \end{bmatrix}; \\
 \Sigma_{10} = \begin{bmatrix} 386.0 \\ -59.2 \quad 366.8 \\ -148.0 \quad -29.8 \quad 406.4 \end{bmatrix}; & \Sigma_{10}^0 = \begin{bmatrix} 386.0 \\ 0 \quad 366.8 \\ 0 \quad 0 \quad 406.4 \end{bmatrix}.
 \end{array}$$

In conjunction with each one of these Σ 's, we carried out three experiments with $T=60, 40, 20$, for a total of 12 experiments. Several algorithms are available in the literature for computing the MP inverse. I used Johnson and Chou's algorithm [5]. As a check, I calculated $\hat{D}^+ \hat{D}$

for several of the \hat{D}^+ we computed. The results were identical to the identity matrix for at least the first six decimals.

The following measures of the relative performance were calculated for the estimated parameters (1) arithmetic mean,² (2) median, (3) standard deviation, (4) root mean square error,² (5) number of the estimates within $\pm 10\%$ of the true parameter, and (6) maximum absolute deviation of the estimates from the parameters. The merits and limitations of several of these measures have been discussed by several authors. See, for example, Summers [9, pp. 12-13]; Quandt [7, pp. 96-97]; and Christ [1, pp. 475-476].

The consistent estimator of σ_{11} provides at the same time what might be viewed as a "non-predictive" measure of the overall goodness of the estimates. As a second measure of the overall goodness of the estimates, I forecasted y_1 at (1) the sample mean value of the exogenous variables and (2) at the sample mean value of the exogenous variable plus one standard deviation. y_2 and y_3 were fixed at their theoretical value for the forecasts. For each experiment I calculated the mean and median of the forecasted y_1 . As a measure of the magnitude of the overall bias of the estimates, I also calculated the norm, $\|\bar{\delta}_{.1}^k - \delta_{.1}\|$, where $\delta_{.1} = (.89 .16 44.00 .74 .13)$, and $\bar{\delta}_{.1}$ is the vector of the average of the estimates derived from the k^{th} estimator. A similar norm was calculated for the median.

Finally, for inferential purposes, one normally needs to attach to the estimate α^* of α a measure of the reliability of the estimate. For LIML and 2SLS, the measure traditionally used is σ_{α}^* , where $[\sigma_{\alpha}^*]^2$ is a consistent estimate of the variance of α^* in the asymptotic distribution of α^* . The idea is that for a relatively large sample, the

distribution of $\sqrt{T}(\alpha^* - \alpha)/\sigma_\alpha^*$ is adequately approximated by a normal distribution with zero mean and unit variance. I calculated $\sqrt{T}(\alpha^* - \alpha)/\sigma_\alpha^*$ for all ILS estimates, and used the Kolmogorov-Smirnov test to test for significant departure from normality. Similar results were calculated for 2SLS and LIML for comparative purposes.

For space limitations, I will not go into the detail of the results, but give a summary of the results in the next section. (Details will be made available to interested readers upon request.)

5. Summary of the Results

Relative Performance

ILS vs. 2SLS: ILS bias tends to be smaller than 2SLS bias. Estimates derived from the two procedures do not appear to differ significantly in concentration or dispersion. The overall goodness of the estimates favors ILS over 2SLS.

ILS vs. LIML: ILS bias tends to be larger than LIML bias. (But the norm of the bias shows the performance evenly divided between the two procedures.) ILS estimates tend to be more concentrated than LIML estimates. The overall goodness of the estimates favors ILS when a non-predictive measure is used; the picture is mixed when a predictive measure is used.²

Effect of Sample Size, Size of Σ and Sparseness of Σ on ILS Bias

The evidence is generally inconsistent with the hypotheses that ILS bias does not depend on the sample size, the size of Σ , and the sparseness of Σ . There is some indication, however, of an interaction between size and sparseness. ILS bias does tend to decrease with sparseness when the size of Σ is large, but not significantly so when the size of Σ is

small to begin with. More research on the interaction between the size and sparseness of Σ may yield useful results.

Effect of Sample Size, Size of Σ and Sparseness of Σ on ILS Relative Performance

The evidence is generally not inconsistent with the hypotheses that the relative performance of ILS does not depend on the sample size, the size and sparseness of Σ .

Tests of the Reliability of ILS Estimates

i) As the sample size increases, the approximation of the distribution of $\sqrt{T}(\alpha^* - \alpha)/\sigma_{\alpha}^*$ by the asymptotic distribution gets better. This is true of the ILS, as well as 2SLS and LIML.

ii) For all cases considered, LIML comes out first in the total as well as in each sample size, followed by ILS and 2SLS. (A similar result was noted by Cragg [2, pp. 101-102] for LIML compared with other consistent estimators (except Full Information Maximum Likelihood (FIML) estimator).)

iii) There is evidence that the adequacy of the normal approximation may depend on the size of Σ . With $T = 20$ or 40 , the approximation appears to work better when the size of Σ is smaller. (Cragg [2, pp. 105-106] found a similar tendency for the "t-ratio" of the consistent estimates he examined.) Generally, the same remarks apply to LIML and 2SLS.

iv) The adequacy of the normal approximation does not appear to be influenced by whether or not the structural disturbances are independent. For ILS (and LIML) this was the case regardless of sample size. For 2SLS, this was the case with $T = 40$ and 60 .

6. Concluding Remarks

A referee thoughtfully observed that the attraction of the proposed method is not so much in its possibly superior small sample behavior, but rather in the possibility of obtaining estimates of the structural parameters without recourse to the data once the reduced form is estimated. The reduced form estimate is all the real world has to give. The structural constraints are the result of theory or intuition. The method proposed "keeps these two sources of 'information' nicely apart. It is one step further on the way to unscrambling the curious mixture of induction and deduction which is so characteristic of applied econometrics." As a follow-up this work will be extended to examine extensively the sensitivity of the estimates to alternative specification of the structural constraints and to deal with other aspects of the Monte Carlo experiments that I have not dealt with at this stage (including the effect of multicollinearity among the exogenous variables).

A second extension relates to the instrumental variable aspect, which I only touched on in this paper. The use of MP inverse opens the way to a family of I.V. estimators in the overidentified case, which have the same structure but differ in the definition of the norm. For example, by premultiplying both sides of (3.15) by $X'X$, we derive the equation of another I.V. estimator (different from ILS) with X as the matrix of instruments. (We have also seen that in the overidentified case 2SLS can be derived by using $X(X'X)^{-1}X'$ instead of the traditional $X(X'X)^{-1}X'Z_1$, as the matrix of instruments, if we are willing to work with the MP inverse.) The behavior of several I.V. estimators in the overidentified case and the question of what constitutes an appropriate norm will be investigated.

Footnotes

¹Recently Swamy and Holmes [10] and Fisher and Wadycki [4] used g-inverse to generalize 2SLS, k class, and 3SLS estimators so that they can be applied to large econometric models when the sample size is smaller than the number of predetermined variables. Unfortunately, the procedure proposed by the authors does not generalize these estimators as claimed. When $G \geq T$ and rank $(X) = T$, these estimators simply do not exist. When X has full row rank, X^- will (minimally) satisfy conditions (i)-(iii) in the text. Its general expression is $X^- = (X'X)^-X'$, (see [8], Theorem 3.2.2, p. 49). It follows that $XX^- = I$, since $X(X'X)^-X'$ is invariant for any choice of $(X'X)^-$. Hence, the general solution $\hat{\Pi}$ for the systematic part of (1.2)—namely $\hat{\Pi} = X^-Y + (I - X^-X)W$, where W is an arbitrary $G \times m$ matrix—is also the general expression for the least squares estimates of the reduced form, which will always satisfy $X\hat{\Pi} = X(X'X)^-X'Y = Y$, when X has full row rank. In the jargon of 2SLS, there is just no way in which the matrix of endogenous explanatory variables that appear in the equation of interest can be purged of its stochastic component, because $\hat{Y} = X\hat{\Pi} \equiv Y$. For a similar reason, the k-class and 3SLS estimators do not exist. The exception occurs when perfect multicollinearity exists among the predetermined variables such that rank $(X) < T$. But this is not the general case of large econometric models, as Fisher and Wadycki [4, p. 463] recognize.

²Some people may question the validity of this measure, unless it is known that the corresponding moment in the population exists. Recent results by Mariano [6] show the 2SLS estimates possess the first two moments for the models we estimated. In light of (2.2) and (3.3), it is reasonable to infer the same is true of the ILS estimates we derived for the same models. The results in the literature do not show LIML possesses a first-order moment. Hence for ILS and 2SLS I used the mean and root mean square error along with the rest of the measures, but for LIML I confined myself to the non-parametric measures.

³It is interesting to note the performance when the covariance is Σ_1 . This is structure 8 taken from Cragg [2] and for which 2SLS and LIML estimates performed well in Cragg's experiments. ILS does better than 2SLS by every summary measure. In comparison with LIML, ILS bias is larger than LIML bias (but the norm of ILS bias is smaller in two out of three cases). ILS estimates tend to be more concentrated around the parameters than LIML estimates. To put these results in perspective, I also compared 2SLS and LIML. 2SLS bias is larger than LIML bias; so is the norm of the bias. Cragg's results showed a slight edge in favor of 2SLS [2, p. 96, experiment 12]. The two measures of concentration and dispersion do not agree on the relative performance of 2SLS vs. LIML.